Computer experiment 1 (Ch2, [Theodoridis 2009])

Notes: In the sequel, it is advisable to use the command

```
randn('seed',0)
```

before generating the data sets, in order to initialize the Gaussian random number generator to 0 (or any other fixed number). This is important for the reproducibility of the results.

Gaussian generator: Generate N l-dimensional vectors from a Gaussian distribution with mean m and covariance matrix S, using the *mvnrnd* MATLAB function.

Solution

Just type

```
mvnrnd(m,S,N)
```

1.

a. Generate a data set X_1 of $N = 1{,}000$ two-dimensional vectors that stem from three equiprobable classes modeled by normal distributions with mean vectors $m_1 = [1, 1]^T$, $m_2 = [12, 8]^T$, $m_3 = [16, 1]^T$ and covariance matrices $S_1 = S_2 = S_3 = 4I$, where I is the 2×2 identity matrix.

b. Apply the Bayesian, the Euclidean, and the Mahalanobis classifiers on X_1.

c. Compute the classification error for each classifier.

2.

a. Generate a data set X_2 of $N = 1{,}000$ two-dimensional vectors that stem from three equiprobable classes modeled by normal distributions with mean vectors $m_1 = [1, 1]^T$, $m_2 = [14, 7]^T$, $m_3 = [16, 1]^T$ and covariance matrices $S_1 = S_2 = S_3 = \begin{bmatrix} 5 & 3 \\ 3 & 4 \end{bmatrix}$.

(b)-(c) Repeat steps b) and (c) of experiment 2.2, for X_2.

3.

a. Generate a data set X_3 of $N = 1{,}000$ two-dimensional vectors that stem from three equiprobable classes modeled by normal distributions with mean vectors $m_1 = [1, 1]^T$, $m_2 = [8, 6]^T$, $m_3 = [13, 1]^T$ and covariance matrices $S_1 = S_2 = S_3 = 6I$, where I is the 2×2 identity matrix.

(b)-(c) Repeat (b) and (c) from experiment 2.2, for X_3.

4.

a. Generate a data set X_4 of $N = 1{,}000$ two-dimensional vectors that stem from three equiprobable classes modeled by normal distributions with mean vectors $m_1 = [1, 1]^T, m_2 = [10, 5]^T, m_3 = [11, 1]^T$ and covariance matrices $S_1 = S_2 = S_3 = \begin{bmatrix} 7 & 4 \\ 4 & 5 \end{bmatrix}$.

(b)–(c) Repeat steps (b) and (c) of experiment 2.2, for X_4.

5.

a. Generate two data sets X_5 and X_5' of $N = 1{,}000$ two-dimensional vectors each that stem from three classes modeled by normal distributions with mean vectors $m_1 = [1,\ 1]^T, m_2 = [4,\ 4]^T, m_3 = [8,\ 1]^T$ and covariance matrices $S_1 = S_2 = S_3 = 2I$. In the generation of X_5, the classes are assumed to be equiprobable, while in the generation of X_5', the *a priori* probabilities of the classes are given by the vector $P = [0.8,\ 0.1,\ 0.1]^T$.

b. Apply the Bayesian and the Euclidean classifiers on both X_5 and X_5'.

c. Compute the classification error for each classifier for both data sets and draw your conclusions.

6.

Consider the data set X_3 (from experiment (2.4)). Using the same settings, generate a data set Z, where the class from which a data vector stems is known. Apply the k nearest neighbor classifier on X_3 for $k = 1$ and $k = 11$ using Z as the training set and draw your conclusions.